FHWA Workshop over the Web for
Travel Model Development
Session 4 Homework
Estimation of a Logit Model

Part 1

Question 1

How do the market shares of modes compare with those in your area?

The auto modes dominate the journey-to-work (JTW) trips. Together, drive alone and the two shared ride modes constitute nearly 86% of the total JTW trips. Transit constitutes nearly 10%, and the non-motorized modes constitute the rest.

This pattern is typical of many urban areas. The transit share can, however, vary quite drastically from 1% to 40% depending on the urban area. Cities like New York and Chicago have had very high transit market shares for the work trips.

Question 2

Comment on the average travel times and costs for each mode.

The average in-vehicle travel times for auto and transit modes are fairly comparable. However, the transit modes have a much higher out-of-vehicle time to account for access, wait, and transfer times. Walk modes, understandably, have the highest travel time.

The costs for the auto modes are the highest for drive alone. For the shared ride modes, the costs are distributed across the trip makers, and are therefore lower. The transit costs are primarily the out-of-pocket costs for fare. The costs vary depending on the mode of payment (monthly passes). Therefore the average fare will be smaller than the fare for a single trip without any concessions.

Question 3

What should the auto costs comprise? Remember, we are talking about out-of-pocket costs for each trip that is made using an automobile.

The costs here include gas, maintenance, and parking. While the gas and maintenance costs are not really incurred for every single trip, they are implicitly accounted for in the trip maker's mode choice decision.

Other costs such as insurance are probably not relevant in a mode choice context because these costs impact the longer-term choice of owning an automobile rather than the short-term choice of whether or not to use the automobile for a work trip.

Part 2

Question 1

What is the dependent variable in this (or any) mode choice model?

Recall that a mode choice model is a "discrete choice" model. It tries to explain the probability of taking each mode for each individual in the sample.

From an estimation standpoint, the dependent variable is the utility associated with each mode. But the utility itself is not observed. What we do observe is the actual choice of the person. The idea behind estimation is to specify the implicit utilities using some observed data (like travel times, costs etc.), then to quantify the probabilities based on the utilities, then maximize the probabilities of the chosen modes.

Question 2

How do you interpret Model 1? How about Model 2?

Model 1 is the most naïve model. Note that the model includes ONLY constants, and also that all the constants are zero. Such a model simply predicts that the market shares of all the modes are exactly the same. Therefore, for this model with six modes, Model 1 predicts that each mode has a share of 16.66%.

Model 2 is also an "all constants" model, but here the constants are not constrained to zero. The constants are actually estimated. This model is popularly called the **market shares model**. This is because the model simply reflects the market shares of the modes in the data sample. In other words, the model does not provide any information about the sensitivity of mode choice to various variables. Instead, it simply tells us what the market shares of the modes are.

How do we estimate the market share of each mode using the market shares model? Refer to Slide 18 in the webinar session. For the market shares model, the deterministic component of the utility consist of ONLY the constant. Therefore, the way to calculate the market share for each mode is as follows:

1. Calculate the exponential of utility for each mode (in this case, the utility is the same as the value of the constant as already discussed).

2. Sum the exponentials across all modes.

3. For each mode, divide the exponential of utility (from step 1) by the sum of exponentials of utilities calculated in step 2.

You will also note that the constants for all modes in Model 2 are negative, except for the drive alone mode. This simply indicates that all things being equal drive alone is the preferred mode. That is, there is a generic bias to the drive alone mode. This is also reflected by the fact that 72% of the work trips are made by driving alone.

Question 3

Model 3 builds on Model 2 by adding an income variable and introducing travel time and cost. Focus on the following questions:

Part a. Comment on the signs of the travel time and cost variables.

The travel time and cost variables are both negative and statistically significant. This indicates that higher the travel time and/or cost of a given mode, the lower the utility of that mode. This is generally true of all level-of-service measures.

Part b. Calculate the implied value of time (VOT) for this model. (Use in-vehicle time for this calculation.)

The implied value of time for this model is calculated as follows:

VOT = (Coefficient of travel time/Coefficient of cost) * (60/100) $/hr

VOT = -0.051 /-0.005 * 0.6 $/hr = 6.12 $/hr

Part c. Is the income variable significant for the auto (Drive alone and shared ride) modes? How do you interpret the significance (or lack thereof)?

We note that the income variable is not significant for the shared ride modes (the absolute value of the t-stat is less than 1.65). This indicates that income has no significant effect on the choice of shared ride over the drive alone mode for work trips.

Part d. Comment on the negative signs of the coefficients of the income variable for the transit, bike, and walk modes. What are these coefficients negative relative to?

The negative signs of the income coefficients on transit, bike and walk modes indicate that all else being equal, higher income individuals prefer to take the drive alone mode over the transit, bike or walk modes.

The negative coefficients are all relative to the drive alone mode.

Question 3

Model 4 builds on Model 3 by separating out motorized and non-motorized travel time variables. It also uses separate coefficients for in-vehicle time and out-of-vehicle time components. Focus on the following questions:

Part a. Notice that the Out-of-Vehicle travel time is not introduced directly. Instead, it is normalized by the trip distance. Why do you think this might be a good idea?

Travelers may be willing to tolerate higher OVTT for longer trips than for shorter trips. To capture this effect of trip distance on sensitivity to out-of-vehicle time, we use OVTT/Distance as a variable in the specification.

Part b. Notice that the travel cost is not introduced directly. Instead, it is normalized by income. Why do you think this might be a good idea?

The sensitivity to travel cost is not the same for all individuals. Typically, higher income individuals have a lower sensitivity to travel cost. To capture the effect of income on the sensitivity to travel cost, we use travel cost/income as a variable in the specification.

Part c. Interpret the coefficients of the autos per worker, CBD Dummy, and Employment density. Note the relative magnitudes and signs relative to the coefficients for other modes.

The autos per worker coefficients indicate that all else being equal, higher the number of autos per worker in a trip maker's household, the lesser the propensity to take any mode other than drive alone.

CBD dummy coefficients indicate the all else being equal, tripmakers are more likely to choose modes other than drive alone if the destination of their trip is in the CBD area. This is intuitive because downtown areas generally have parking limitations.

The employment density coefficients indicate that all else being equal, tripmakers are less likely to choose drive alone for destinations in dense employment areas. This variable may be capturing some of the same effects that the CBD dummy is capturing.

Part d. Compute the value of motorized in-vehicle time. Be careful with the cost coefficient here. It is normalized by income!

The value of motorized in-vehicle time can be computed as follows:

$$\text{VOT} = \frac{\beta_{\text{IVTT}}}{\left(\beta_{\text{COST/INC}} / \text{Inc}\right)} * 60/100$$

= -0.02*0.6/(-0.052/42.4) = 9.8 $/hr

Here, we assumed the average income to compute the value of time. The average income was assumed to be $42, 400 per year.

Question 3

Look at the log-likelihood value at convergence (Row 50), and the rho-squared values (Rows 51 and 52). What do you observe? What does this tell about the relative performance of the models?

We note that the log-likelihood at convergence becomes increasingly positive (or less negative) as we move from model 1 to model 4. We also note that the rho-squared values gradually increase from model 1 through model 4. These statistics indicate that each model is progressively better at explaining the variability in mode choice.